

Analisis Sentimen Sistem E-Tilang Menggunakan Algoritma Naive Bayes Dengan Optimalisasi Information Gain

Rakhmi Khalida^{1,*}, Siti Setiawati¹

¹ Teknik Informatika; Universitas Bhayangkara Jakarta Raya; Jl Raya Perjuangan No.81 Bekasi Utara, (021) 88955882; e-mail: rakhmi.khalida@dsn.ubharajaya.ac.id; siti.setiawati@dsn.ubharajaya.ac.id

* Korespondensi: e-mail: rakhmi.khalida@dsn.ubharajaya.ac.id

Diterima: 8 Mei 2020; Review: 9 Mei 2020; Disetujui: 27 Mei 2020; Diterbitkan: 17 Juni 2020

Abstract

The Government of Indonesia took steps to change the system to improve public services in traffic violations by implementing the e-ticketing system. This system is a solution for disciplining motorized motorists from committing traffic violations. The existence of e-ticketing is also a solution to prevent the delinquency of law enforcers from illegal levies, peace terms in place, to accountability of fines. In this study, sentiment analysis of the e-ticketing system or opinion mining to classify the variety of public comments that give a positive, negative or neutral impression. Twitter social media is one of the objects to express opinions because it is user friendly, updated topics, and openly accesses tweets. Opinions on Twitter are collected, then the preprocessing stage is performed, then the selection of information gain features helps reduce noise caused by irrelevant labels, the next step is the classification of sentiments with the Naïve Bayes algorithm and finally polarity sentiments. This research resulted in an accuracy of 41.82%, a precision of 50.51% and a recall of 45.45%.

Keywords: Sentiment analysis, E-ticketing, Information Gain, Naive Bayes

Abstrak

Pemerintah Indonesia melakukan langkah perubahan untuk memperbaiki sistem pelayanan publik dalam pelanggaran berlalu-lintas yaitu dengan menerapkan sistem e-Tilang. Sistem ini menjadi solusi mendisiplinkan para pengendara kendaraan bermotor dari banyaknya melakukan pelanggaran berlalu-lintas. Keberadaan e-Tilang juga menjadi solusi mencegah kenakalan penegak hukum dari pungutan liar, istilah damai ditempat, hingga akuntabilitas uang denda. Dalam penelitian ini melakukan analisis sentimen tentang sistem e-Tilang atau opinion mining untuk mengelompokan ragam komentar masyarakat yang memberikan kesan positif, negatif atau netral. Media sosial Twitter menjadi salah satu objek untuk menyampaikan opini karena user friendly, topik ter-update, dan terbuka mengakses tweet. Opini pada twitter dikumpulkan, lalu dilakukan tahapan preprocessing, selanjutnya dengan seleksi fitur information gain membantu mengurangi noise yang disebabkan oleh label-label yang tidak relevan, tahap selanjutnya adalah klasifikasi sentimen dengan algoritma Naïve Bayes dan terakhir sentimen polarity. Penelitian ini menghasilkan accuracy 41,82%, presisi 50,51% dan recall 45,45%.

Kata kunci: Analisis sentimen, E-Tilang, Information Gain, Naive Bayes

1. Pendahuluan

Penerapan e-Tilang merupakan langkah baru yang diambil pemerintah dalam mewujudkan Good Governance agar pelayanan publik yang diberikan dapat lebih efektif, efisien, transparan dan akuntabel. e-Tilang merupakan digitalisasi proses tilang, dengan

memanfaatkan teknologi diharapkan seluruh proses tilang akan lebih efisien (Rakhmadani, 2017). Pasal 272 Undang – Undang No. 22 Tahun 2009 tentang Lalu Lintas Jalan dan Angkutan Jalan menyebutkan bahwa untuk mendukung kegiatan penindakan pelanggaran di bidang lalu lintas dan angkutan jalan, dapat digunakan peralatan elektronik. Hasil penggunaan alat elektronik ini dapat digunakan sebagai bukti di pengadilan. Peralatan elektronik adalah alat perekam kejadian untuk menyimpan informasi (Pasal 272 UU No. 22, 2009). Pada pasal tersebut diatur lebih lanjut dalam Peraturan Pemerintah No 80 Tahun 2012 tentang Tata Cara Pemeriksaan Kendaraan Bermotor di Jalan Dan Penindakan Pelanggaran Lalu Lintas dan Angkutan Jalan, yang mengatur bahwa penindakan Pelanggaran Lalu Lintas dan Angkutan Jalan di dasarkan atas hasil : (1) Temuan dalam proses pemeriksaan kendaraan; (2) laporan; dan/atau (3) rekaman peralatan elektronik (PP No.80, 2012).

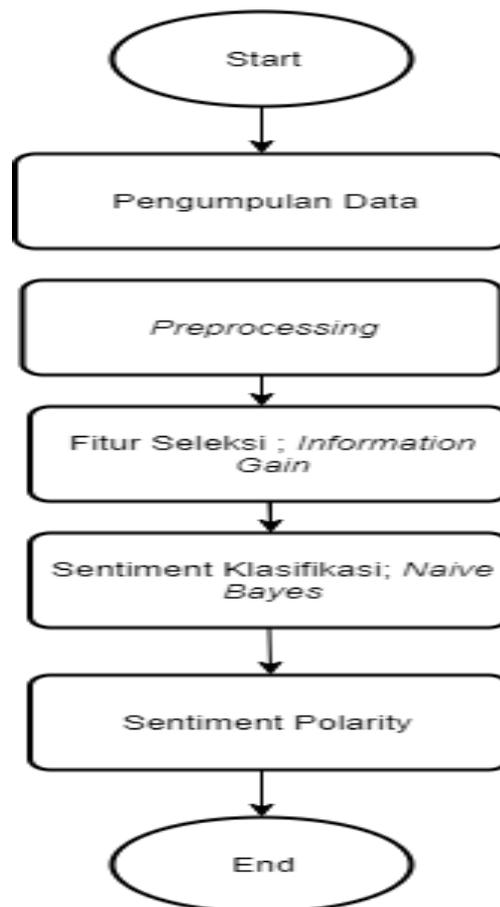
Sistem e-Tilang menjadi solusi mendisiplinkan para pengemudi kendaraan bermotor dari banyaknya melakukan pelanggaran berlalu-lintas. Keberadaan e-Tilang juga menjadi solusi mencegah kenakalan penegak hukum dari pungutan liar, istilah damai ditempat, hingga akuntabilitas uang denda. Efektif dan efisien sistem e-Tilang menimbulkan ragam komentar pada masyarakat. Perkembangan website 4.0 dan tingginya penggunaan media sosial dapat dimanfaatkan untuk menjadi sumber informasi dan pengambilan keputusan, dengan melakukan text mining yang bersumber dari komentar pada media sosial ataupun website dapat melakukan analisis sentimen. Text mining merupakan salah satu teknik yang digunakan untuk menggali kumpulan dokumen text sehingga dapat diambil intisaryanya (Budi, 2017) Penelitian yang telah dilakukan oleh Indrayuni analisis sentimen atau opinion mining merupakan salah satu solusi mengatasi masalah untuk mengelompokkan opini atau review menjadi opini positif atau negatif atau netral secara otomatis (Indrayuni, 2016). Analisis sentimen dilakukan untuk mendapatkan informasi yang dapat digunakan sebagai manajemen reputasi dan sarana evaluasi bagi perusahaan (Oktasari, 2016). Hal yang dilakukan dalam penelitian ini adalah melakukan text mining pada postingan mengenai sistem e-Tilang pada media sosial Twitter menggunakan algoritma klasifikasi Naive Bayes dan dioptimalisasi dengan seleksi fitur information gain. Penelitian ini membantu pemerintah dalam melakukan evaluasi program yang dijalankan, memberikan informasi sentiment masyarakat terhadap sistem e-Tilang. Tanpa adanya evaluasi dari masyarakat, pemerintah tidak akan mengetahui keberhasilan sistem yang dijalankan.

Terdapat beberapa penelitian yang dilakukan mengenai analisis sentimen diantaranya oleh Mahendrajaya, Buntoro, dan Setyawan melakukan analisis sentimen pengguna gopay menggunakan metode lexicon based dan support vector machine menghasilkan pelabelan dengan Lexicon Based berjumlah 923 untuk positif dan 287 untuk negatif. Sedangkan klasifikasi metode SVM menggunakan kernel Linear menghasilkan 89,17% dan 84,38% untuk kernel Polynomial (Mahendrajaya, 2019). Penelitian oleh Ruhana melakukan analisis sentimen terhadap penerapan sistem plat nomor ganjil/genap pada twitter dengan metode klasifikasi naive bayes menghasilkan klasifikasi teks dalam bentuk positif dan negatif untuk penerapan lalu lintas ganjil genap, dalam penelitian ini menghasilkan accuracy 86,67%, precision 71,43% dan

recall 80,00% (Ruhana, 2019). Algoritma klasifikasi yang dapat melakukan teks mining diantaranya Support Vector Machine (SVM), Naïve Bayesian classification (NBC) dan K-Nearest Neighbor (K-NN) (Ipmawati, 2017).

2. Metode Penelitian

Rancangan penelitian yang akan dilakukan dibagi menjadi beberapa bagian, yaitu pengumpulan data, preprocessing, fitur seleksi dengan information gain, sentimen klasifikasi dengan Naive Bayes dan sentimen polarity. Perangkat lunak yang digunakan untuk melakukan analisis sentimen yaitu Rapidminer 9.5.



Sumber: Hasil Penelitian (2020)

Gambar 1. Rancangan Penelitian

2.1. Pengumpulan Data

Data diambil dengan metode crawling dari media sosial Twitter menggunakan bantuan API key. Pada aplikasi rapidminer melalui API dengan query e-Tilang. Setelah mendapatkan dataset dari twitter setelah itu export ke format excel.

2.2. Preprocessing Data

Preprocessing dilakukan untuk menghindari data yang kurang sempurna, gangguan pada data, dan data-data yang tidak konsisten (Hemalatha, 2012). Tahapan preprocessing data diantaranya:

1. Tokenisasi

Tokenisasi yaitu mengumpulkan semua kata yang muncul dan menghilangkan baca maupun simbol apa pun yang bukan huruf (Muthia, 2017)

2. *Stemming*

Stemming adalah tahap mencari akar kata dengan menghilangkan imbuhan pada sebuah kata (Fedlman, 2013).

3. *Stopword*

Stopword merupakan kata-kata yang tidak memiliki pengaruh terhadap proses klasifikasi.

2.3. Fitur Seleksi

Fitur seleksi yang digunakan dalam penelitian ini adalah information gain. Information Gain merupakan metode seleksi fitur paling sederhana dengan melakukan perangkingan atribut dan banyak digunakan dalam aplikasi kategorisasi teks, analisis data micro array dan analisis data citra. Information Gain dapat membantu mengurangi noise yang disebabkan oleh label-label yang tidak relevan. Information Gain mendeteksi fitur-fitur yang paling banyak memiliki informasi berdasarkan kelas tertentu.

Information Gain merupakan teknik seleksi fitur yang memakai metode scoring untuk nominal ataupun pembobotan atribut kontinu yang didiskretkan menggunakan maksimal entropy. Suatu entropy digunakan untuk mendefinisikan nilai Information Gain. Entropy menggambarkan banyaknya informasi yang dibutuhkan untuk mengkodekan suatu kelas. Information Gain (IG) dari suatu term diukur dengan menghitung jumlah bit informasi yang diambil dari prediksi kategori dengan ada atau tidaknya term dalam suatu dokumen. Secara matematis dituliskan dengan (Indah, 2016).

$$\text{InfoGain}(S, A) = \text{Entropy}(S) - \sum_{v \in \text{Value}(A)} \frac{|S_v|}{S} \text{Entropy}(S_v) \tag{1}$$

$$\text{Entropy}(S) = - \sum \frac{|S_i|}{S} \log \frac{|S_i|}{S} \tag{2}$$

Dimana S adalah jumlah seluruh fitur, A adalah kategori, Sv adalah jumlah sampel untuk nilai v, v adalah nilai yang mungkin untuk kategori A, Si adalah fitur ke i dan Value(A) adalah himpunan nilai-nilai yang mungkin untuk kategori A. Fitur yang dipilih adalah fitur dengan nilai Information Gain yang tidak sama dengan nol dan lebih besar dari suatu nilai threshold tertentu. Ide dibalik Information Gain untuk memilih fitur adalah menyatakan fitur dengan informasi yang paling signifikan terhadap kategori (Indah, 2016).

2.4. Naïve Bayes

Naive Bayes Classifier (NBC) merupakan metode yang berdasarkan atas probabilitas bayes untuk melakukan pengelompokan data. Tahapan Algoritma Naive Bayes:

Hitung probabilitas bersyarat :

$$P(x | C) = P(x_1, x_2, \dots, x_n | C) \tag{1}$$

C = class

x = vektor dari nilai atribut n

$P(x_i|C)$ = proporsi dokumen dari class C yang mengandung nilai atribut x_i .

Hitung probabilitas prior untuk tiap class:

$$P(C) = N_j/N \quad (2)$$

N_j = jumlah dokumen pada suatu class

N = jumlah total dokumen

Hitung probabilitas posterior dengan rumus:

$$P(C|x) = (p(x|C)p(c))/(p(x)) \quad (3)$$

3. Hasil dan Pembahasan

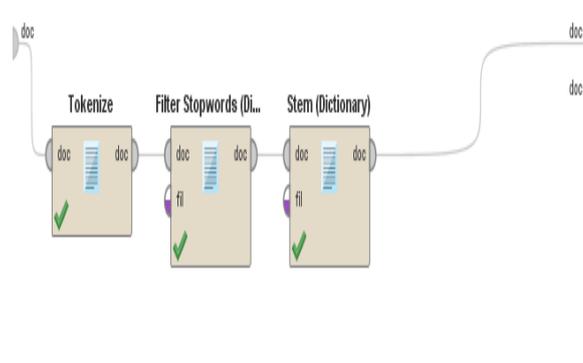
Dataset diambil dari media sosial twitter selama 7 hari terakhir dengan jumlah data sebanyak 276 data.

Tabel 1. Sampel Dataset

No	Ulasan	Nilai
1	pasti kota lain akan menyusul! dan aku senang makin sedikit yg melanggar lalu lintas karena takut kena e tilang :) https://t.co/DZERHZImFF	Positif
2	@DesaKita2 Ah jancuk kok, aku ming khilaf sitik wae di tilang. Pokokke polisine jeneng e eko	Negatif
3	RT @AmadRowi: @e100ss Contoh surat hasil E Tilang/E-TLE Tapi bukan milik saya yaaa https://t.co/aljFn5WTe9	Netral
4	ayo surabayans luwih taat peraturan lalu lintas ya ben ga kene tilang!?? https://t.co/ifXKpjMH9n	Positif
5	E-Tilang Surabaya, Pencuri Bisa Dideteksi Hanya Dari Bajunya #ETLE #Etilang #RepublikMerdeka https://t.co/Y66riAI4TE	Netral
6	Akhir Januari 2020, Pengemudi Motor di Jalan Sudirman-MH Thamrin Bisa Kena E-Tilang https://t.co/z5BFLzu1ys	Netral
7	Kakorlantas Ingin Sistem E-tilang Dicontoh Seluruh Polda https://t.co/FeeA0nDYCf https://t.co/PVwqyvVvBz	Positif
8	Surabaya Terapkan E-Tilang, Risma: Ada Kamera Tiap 15 Meter https://t.co/uxkrlH63P6	Netral
9	@rosihanazwar @JJRizal Sebenarnya sdh diinisiasi secara nasional ya sejak 2015, ujicoba, sampai sekarang resmi sejak Juli 2019. Penindakan e tilang dari bukti rekaman cctv, juga harus segera diproses sesaat setelah kejadian dari rekaman yg "terpantau" . Jd tetap tergantung yg memantau...	Negatif
10	Warga Apresiasi Sat Lantas Polres Simalungun Bantu Mengurus Pembayaran Denda E-Tilang https://t.co/OtrEX395iA	Positif

Sumber: Hasil Penelitian (2020)

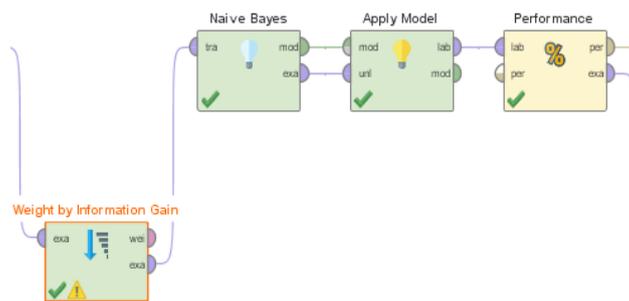
Data yang telah terkumpul, seperti pada tabel 1 akan melalui tahap preprocessing seperti pada gambar 2.



Sumber: Hasil Penelitian (2020)

Gambar 2. Tahapan Preprocessing

Tahapan selanjutnya adalah seleksi fitur menggunakan information gain, lalu tahap selanjutnya melakukan sentimen klasifikasi menggunakan algoritma Naïve Bayes seperti gambar 3.

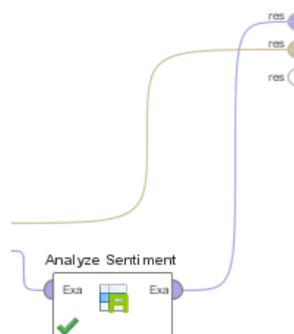


Sumber: Hasil Penelitian (2020)

Gambar 3. Tahapan Fitur Seleksi dan Sentimen klasifikasi

3.1. Analisis Sentimen

Noise yang telah hilang dan data yang telah bersih, selanjutnya melakukan tahapan analisis sentimen, seperti pada gambar 4.



Sumber: Hasil Penelitian (2020)

Gambar 4. Tahapan Analisis Sentimen

Dalam penelitian ini algoritma menghasilkan akurasi 41,82% dengan menggunakan metode Naive Bayes dari pengujian data yang telah dikumpulkan dan dengan diuji dengan data training dan data testing. Dari pengujian hasil data tersebut dengan rapidminer akan mendapatkan class precesion dan class recall, yang dimaksud dengan precesion yaitu perbandingan dokumen yang relevan terhadap total dokumen hasil query sedangkan recall perbandingan jumlah dokumen relevan terambil terhadap total dokumen relevan (Mahendrajaya, 2019). Hasil pengujian dapat dilihat pada tabel 2.

Tabel 2. Hasil Pengujian

	True Positive	True Neutral	True Negative	Class Precission
Pred Positive	100	42	56	50,51%
Pred Neutral	48	90	78	41,67%
Pred Negative	72	88	86	34,96%
Class Recall	45,45%	40,91%	39,09%	

Sumber: Hasil Penelitian (2020)

Perhitungan manualnya adalah sebagai berikut:

Nilai akurasi =

$$(100+90+86)/(100+48+72+42+90+88+56+78+86) \quad (4)$$

$$= 276/660=41,82 \%$$

Nilai presisi positif =

$$100/(100+42+56) \quad (5)$$

$$= 100/198=50,51 \%$$

Nilai presisi negatif =

$$(86)/(72+86+88) \quad (6)$$

$$= 86/246=34,96 \%$$

Nilai presisi netral =

$$(90)/(48+90+78) \quad (7)$$

$$= 90/216=34,96 \%$$

Nilai recall positif =

$$(100)/(100+48+72) \quad (8)$$

$$= 276/220=45,45 \%$$

Nilai recall negatif =

$$(86)/(56+78+86) \quad (9)$$

$$= 86/220=39,09 \%$$

Nilai recall netral =

$$90/(42+90+88) \quad (10)$$

$$= 90/220=40,91 \%$$

4. Kesimpulan

Penerapan fitur seleksi dalam pengoptimalisasi tingkat akurasi analisis sentiment sistem e-Tilang telah dapat dilakukan dan telah dapat memberikan sebuah peningkatan akurasi terhadap model Naive Bayes yang merupakan model klasifikasi dan memberikan hasil penelitian dengan hasil accuracy 41,82%, presisi 50,51% dan recall 45,45%.

Daftar Pustaka

- Budi, S. (2017), "Text Mining Untuk Analisis Sentimen Review Film Menggunakan Algoritma K-Mean,". Techno.Com
- Feldman, R. (2013). "Techniques and applications for sentiment analysis," Communications of the ACM, vol 56, no 4, pp.82,
- Hemalatha, I., Varma, D. G. P. S., dan A. Govardhan, D, (2012) "Preprocessing The Informal Data for Efficient Sentiment Analysis," International Journal of Emerging Trends & Technology in Computer Science, Vol 2, no 58.
- Indah Maulida, Addy Suyatno, Heliza Rahmania Hatta, (2016). "Seleksi Fitur Pada Dokumen Abstrak Teks Bahasa Indonesia Menggunakan Metode Information Gain" JSM, Vol 17 No.2.
- Indrayuni, E, "Analisa Sentimen Review Hotel Menggunakan Algoritma Support Vector Machine Berbasis Particle Swarm Optimization," Jurnal Evolusi, Volume 4, 2016.
- Ipmawati, J., Kusriani, dan Taufiq Luthfi, E. (2017). "Komparasi Teknik Klasifikasi Teks Mining Pada Analisis Sentimen". In Indonesian Journal on Networking and Security, Vol. 6.
- Mahendrajaya S., Buntoro G.A dan Setyawan M.B. (2019). "Analisis Sentimen Pengguna Gopay Menggunakan Metode Lexicon Based dan Support Vector Machine," Komputek Jurnal Teknik Universitas Muhammadiyah Ponorogo, Vol 3 no 2, pp. 52-63,
- Muthia, D. A. (2017). "Analisis Sentimen Pada Review Restoran Dengan Teks Bahasa Indonesia Menggunakan Algoritma Naive Bayes," Jurnal Ilmu Pengetahuan Dan Teknologi Komputer, vol 2, no 2, p. 39-45.
- Pasal 272 Undang – Undang No. 22 Tahun 2009 tentang Lalu Lintas Jalan dan Angkutan Jalan. Peraturan Pemerintah No 80 Tahun 2012 tentang Tata Cara Pemeriksaan Kendaraan Bermotor di Jalan Dan Penindakan Pelanggaran Lalu Lintas dan Angkutan Jalan.
- Oktasari, L., Chrisnanto, Y. H., dan Yuniarti, R, (2016). "Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naive Bayes Classifier," Prosiding SNST.
- Rakhmadani S. (2017). "Analisis Penerapan E-Tilang Dalam Mewujudkan Good Governance di Indonesia," Prosiding SNaPP2017.
- Ruhyana, N. (2019). "Analisis Sentimen Terhadap Penerapan Sistem Plat Nomor Ganjil / Genap Pada Twitter Dengan Metode Naive Bayes," Jurnal IKRA-ITH Informatika, Vol 3, no 1, pp.94-99.